CS 134
Problem Set 3

Xiner Zhou

February 18, 2017

1. Non-Preferential Attachment (20 points).

a. (8 points)

Let $Z_k(t)$ be the random variable denoting node i_k receives an edge at time t, where $k \in \{1, \ldots, t-1\}$. For $t > k$:

$$Z_k(k+1) = \frac{1}{k}$$

$$Z_k(k+2) = \frac{1}{k+1}$$

$$\cdots$$

$$Z_k(t) = \frac{1}{t-1}$$

Let $X_k(t)$ be the random variable denoting the in-degree of node i_k at time t, where $k \in \{1, \ldots, t-1\}$. Then,

$$X_k(t) := \sum_{T=k+1}^{t} Z_k(T)$$

$$= \frac{1}{k} + \frac{1}{k+1} + \cdots + \frac{1}{t-1}$$

$$= \sum_{i=1}^{t-1} \frac{1}{i} - \sum_{i=1}^{k-1} \frac{1}{i}$$

$$\approx log(t-1) - log(k-1)$$

$$= log(\frac{t-1}{k-1})$$

b. (8 points)

$$X_k(t) = log(\frac{t-1}{k-1}) \geq d \tag{1}$$

$$\Rightarrow \frac{t-1}{k-1} \geq e^d \tag{2}$$

$$\Rightarrow k \leq \frac{t-1}{e^d} + 1 \tag{3}$$

Therefore, at time step t, the fraction of nodes that have expected in-degree at least d is:

$$\frac{\frac{t-1}{e^d} + 1}{t} = \Theta(\frac{1}{e^d}) = \Theta(e^{-d})$$

c. (4 points)

No, I don't expect the graphs genreated using this method to have a power-law degree distribution.

If we believe the rich-get-richer model or the preferential-attachment model explains the underlying mechanism of the observed "popularity" in network follows a power-law distribution, then, we have proven in class that the number of in-degree will be distributed approximated as a power-law $ck^{-\alpha}$, with $c = \frac{1-p}{p}$ and $\alpha = \frac{1}{1-p}$ where p is the probability of a node choosing existing a node uniformly at random from all previous nodes. Power law arises from the feedback introduced by correlated decisions across population. In the rich-get-richer model, this dependency or correlation is represented by $1 - p$: as p gets smaller, $1 - p$ gets larger, dependency is stronger, that is, copying becomes more frequent, making it more likely to see extreme popular items in the network, this is when the power law is stronger.

In this case, we have $1 - p = 0$, basically there is no copying mechanism or no preferential attachment, there is an absence of fundamental mechanism that generates power-law phenomena.

2. Power Laws in Classrooms [Easley and Kleinberg, 18.8 Q3] (10 points). I expect that Harvard classes to be more closely follow a power-law distribution as a function of k.

Power-law distribution is widespread in cases where the quantity being measured can be viewed as a type of popularity (this case certainly is!). And we use the rich-get-richer model or preferential-attachment model to explain the mechanism why the power law arises. The essence of the rich-get-richer model is the copying mechanism, which basically saying: every node in the network has a probability of p not to copy the decision of a random previous people; and has a probability of $1 - p$ to copy previous people's choices, and end up linking to some node i with the probability proportional to the number of links i already has. Because the probability of node i increases popularity is directly proportional to the its current popularity. This so-called preferential-attachment phenomena is the basic mechanism that gives rise to power-law distribution, and is stronger when $1 - p$ is larger, that is, when people tend to copy other's behavior more often.

Harvard student group is a much smaller, more inter-connected social network, compared to the third-grade school students spread out through New York state, where kids in different districts may not even know each other. Beside the connectivity structure of the two networks, Harvard students are more likely to seek class information and based their decision on information through peers. College classes is more challenging at Harvardl, and which classes to take partial decide the college experience, as well as future career. Therefore, Harvard college students have more incentives to share and seek class information, and they have more resources to help them do so, like reviews and ratings from previous takers.

Therefore, the likelihood of preferential attachment is much higher for Harvard classes, and thus is more likely to follow a power-law distribution.

3. It's a Power-Law Story (15 points). First, calculate the value of c:

$$\sum_{k=1}^{485,165,195} P(X = k) = c \sum_{k=1}^{485,165,195} \frac{1}{k^2} = 1$$

$$\Rightarrow c = \frac{1}{\sum_{k=1}^{485,165,195} \frac{1}{k^2}} \approx \frac{6}{\pi^2}$$

2

Second, calculate the expected rerevue from selling the single:

$$E(Revenu) = \sum_{k=1}^{485,165,195} P(X = k) \times k$$

$$= \sum_{k=1}^{485,165,195} c\frac{1}{k^2} \times k$$

$$= c \sum_{k=1}^{485,165,195} \frac{1}{k}$$

$$= \frac{6}{\pi^2}log(485, 165, 195)$$
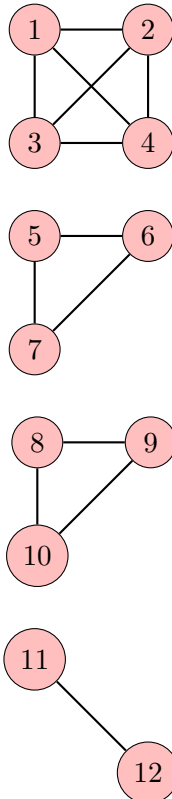
$$= \frac{6}{\pi^2} \times 20$$

$$= 12.15854$$

Finally, calculate the expected profit:

$$E(Profit) = E(Revenu) - 100 = -87.84146 < 0$$

Therefore, people should not purchase the right, if trying to make a positive profit.

4. Power-Law Cliques (25 points).

a. (3 points)

b. (7 points)

For every node $i \in V$ with degree k, the clique generated C_i contains the node i iteself plus all its neighbors, and they are pairwise connected. Therefore, there are $k + 1$ nodes in the clique C_i, and every node has degree of k. Run through all nodes in G, we get the degree distribution of $\widetilde{G}$:

$$\widetilde{p}(x = k) = \frac{Number of nodes with degree k}{Total number of nodes in \widetilde{G}}$$

$$= \frac{|V| \times p(x = k) \times (k + 1)}{\sum_{k=0}^{\triangle} |V| \times p(x = k) \times (k + 1)}$$

$$= \frac{p(x = k) \times (k + 1)}{\sum_{k=0}^{\triangle} p(x = k) \times (k + 1)}$$

c. (8 points)

Suppose that p has the power-law property until $\triangle$, that is:

$$\frac{p(k\prime)}{p(k)} \geq \left(\frac{k\prime}{k}\right)^{-\beta}$$

$$\Rightarrow \frac{\widetilde{p}(k\prime)}{\widetilde{p}(k)} = \frac{\widetilde{p}(k\prime)(k\prime + 1)}{\widetilde{p}(k)(k + 1)}$$

$$\geq \left(\frac{k\prime}{k}\right)^{-\beta} \frac{k\prime + 1}{k + 1} \geq \left(\frac{k\prime}{k}\right)^{-(\beta+1)}$$

The last inequality is due to the fact that $\left(\frac{k\prime+1}{k+1}\right)^{-1} \geq \left(\frac{k\prime}{k}\right)^{-1}$.

Therefore, $\widetilde{p}$ has the power-law property until $\triangle$ as well with constant $\beta + 1$.

d. (7 points) In networks whose degree distributions have the power law property, the "strong form" of the friendship paradox does not necessarily hold true. We have seen above that the graph $\widetilde{G}$ has a power law property, but for each node i in the network, the avergae degree of $i's$ neighbors is equal to $i's$ own degree, since all i's neighbors have the same degree as i does. This statement is subtly different from what we talked about in class. in the sense that, the "friendship paradox" focuses on the average degree of all nodes and avergae degree of all the neighbors, over all nodes in the network; while here we are talking about a specific node i, not average across the network.

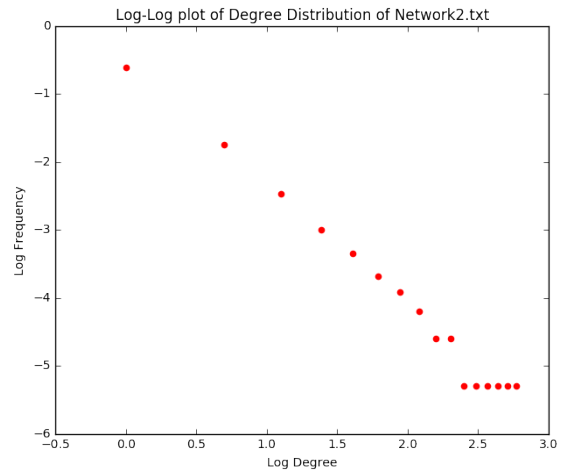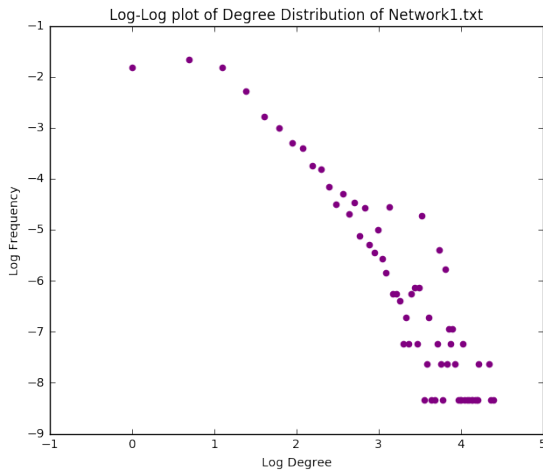**5. Coding: Fitting Power-Laws (30 points).**

a. (5 points)

The degree distribution of network1 and network2, respectively.
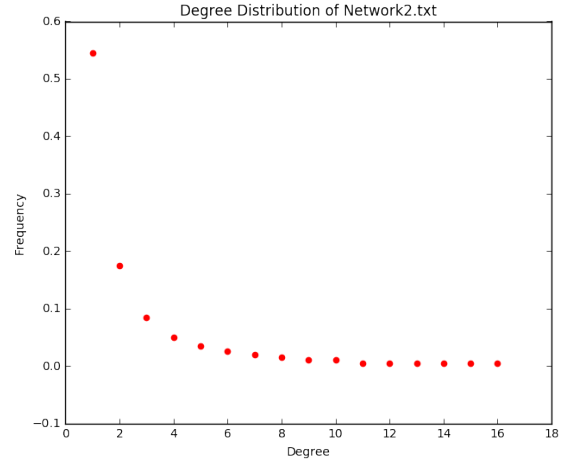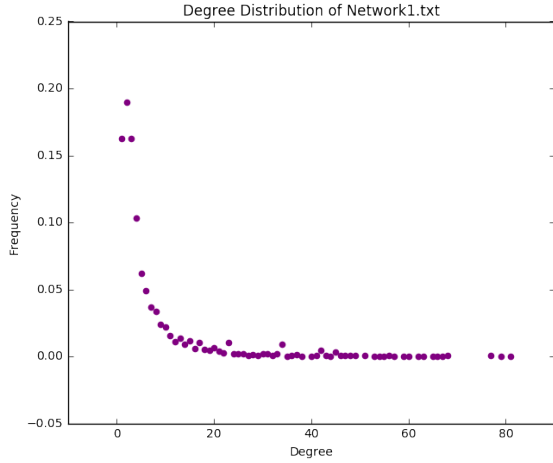
b. (5 points)

The log-log plot of degree distribution of network1 and network2, respectively.

c. (7 points)

Network 1 has α estimated using OLS: 1.9741.

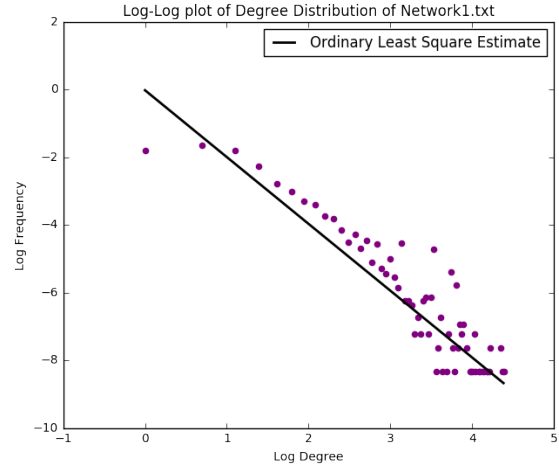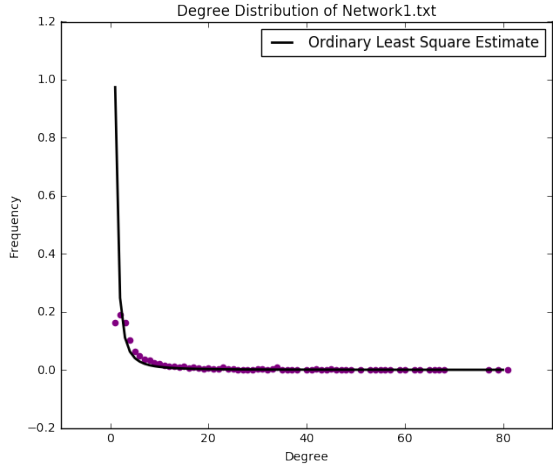Network 2 has α estimated using OLS: 1.8204.

d. (10 points)

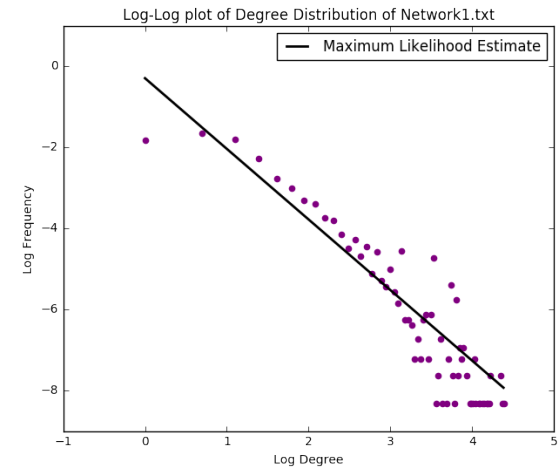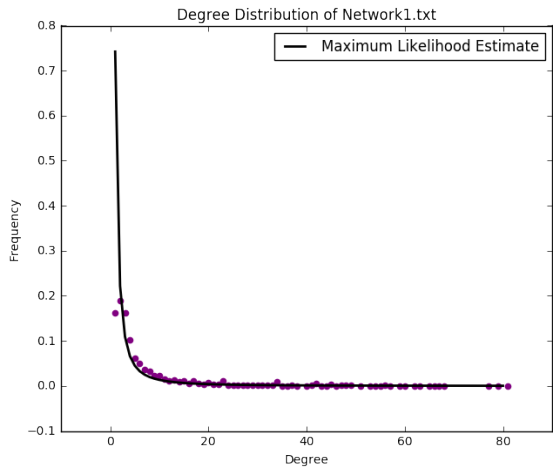Network 1 has α estimated using Maximum Likelihood: 1.7422.

Network 2 has α estimated using Maximum Likelihood: 2.7299.
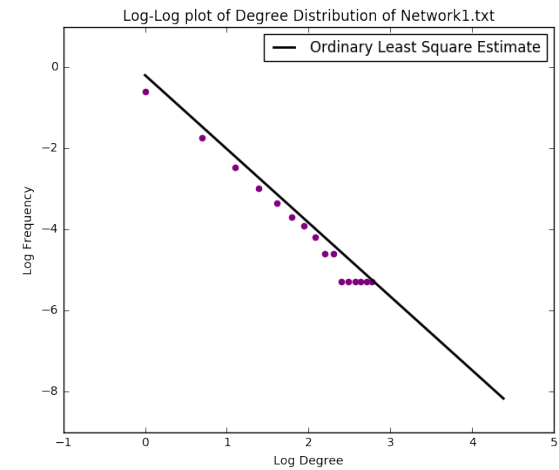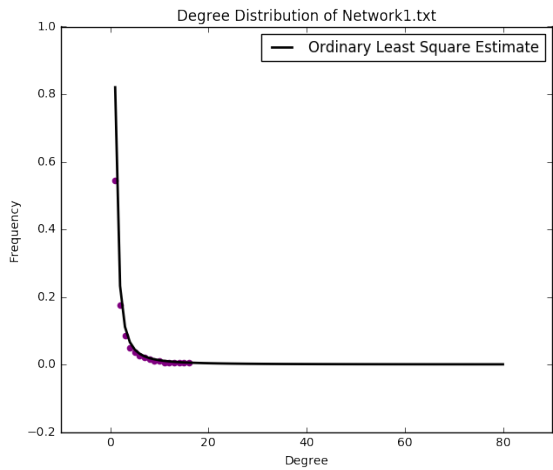
e. (3 points)

For Network1, using OLS, degree distribution and log-log plot of degree distribution, respectively.
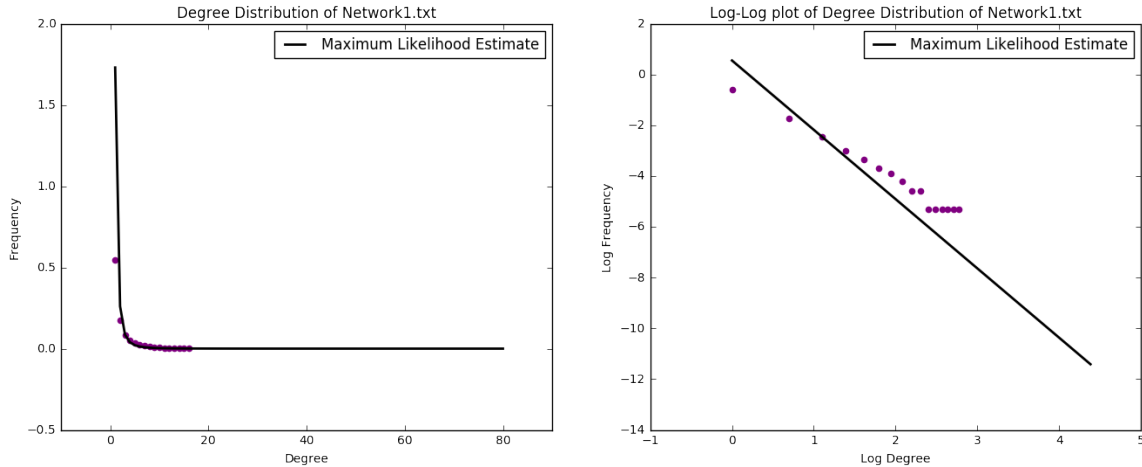
For Network1, using MLE, degree distribution and log-log plot of degree distribution, respectively.

For Network2, using OLS, degree distribution and log-log plot of degree distribution, respectively.

For Network2, using MLE, degree distribution and log-log plot of degree distribution, respectively.

The Ordinary Least Square approach has one major drawback, that is: we try to fit a straight line with 2 degrees of freedom (the intercept and the slope), assuming that the estimated slope approximates the α of our power-law distributino, while allowing the intercept estimated separately to minimize the sum of the squared residuals. However, if the data follows a power-law distribution as we derived in class, then the log-log plot's intercept is determined by the α as well, in other words, the intercept and the slope should not be estimated separately, as what OLS does. We actually only have one parameter alpha to estimate, not two parameters, therefore, technically, it does not satisfy ordinary least square approach assumption. Although convenient to estimate via OLS, the estimated α is not accurate.

The maximum likelihood is more appropriate, in the sense that, given the assumtption that the data follows a power-law distribution, we don't put additional unrealistic assumptions into the estimation process. We simply ask the parameter α for which the observed data is most likely to been drawn from the distribution. Therefore, we should estimate via ML, rather than OLS approach.